

GEOG0013 Data Analysis Project Worksheet

Before you begin

To complete the worksheet you will need to make use of the datasets, techniques and R code you have worked with on the guided RStudio practicals in Week 9 which are available on Moodle. The worksheet can be completed as you work through the practicals OR you can go through the guidance notes in the practicals first and then complete the worksheet – it is up to you.

Background

You are tasked with creating a portrait of Inner London's neighbourhoods (electoral wards) using the secondary datasets provided for the Week 9 practical exercise. These are available on Moodle. You have access to two sets of files:

1. A bundle of files starting with the prefix '**InnerLondon_ward_polygons**'. These make up a vector *shapefile* of Inner London wards.
2. A .csv file called **Wards_data.csv** containing statistical data about each Inner London ward. The variables in this file are as follows:

Variable name	Contents
Ward_code	A unique code identifying each ward
Ward_name	Name of the ward
Borough	Borough the ward is located within
over65	% of ward population aged 65+
wbritish	% of ward population who report their ethnicity as White British
bornuk	% of ward population born in the UK
loneparent	% of ward households that are lone parent households
lifeexp	Female life expectancy at birth
liti	% of ward population with a long-term limiting illness/disability
degree	% of ward adults with degree level qualifications
unemp	% of working age population who are unemployed
kidspoverty	% of children living in poverty
income	Median ward household income in £1000
nocar	% of ward households with no car or van
biketowork	% of employed population cycling to work
onebeds	% of dwellings with one bedroom
rented	% of households renting their dwelling
houseprice	Median ward house price in £1000
crimerate	Reported crimes per thousand population
turnout	Voter turnout in London mayoral election

The variables were obtained from the London Datastore ward profile database. They come from different raw sources and were measured at slightly different time points but you do not need to worry about that in this exercise.

Your tasks

To complete these tasks you will need to adapt the R code learnt in the practicals.

1. Pick *one dependent variable* (y) to focus on from those available in the .csv file that you are interested in. **Do not pick unemployment as this was used in the practical exercise!**
2. Using appropriate techniques, describe how the values on your selected y variable vary across Inner London. *Hint:* there are lots of ways you could do this – statistics, plots and maps could all work and the choice is up to you.
3. Select *two additional independent variables* (x variables) you think could be associated with your dependent y variable. Examine the relationships between the x variables and your y variable using appropriate (i) correlation tests and (ii) regression models as well as graphs if you want. **Do not look at the relationship between unemployment and income as this was already covered in the practicals!** Interpret your results. *Hint:* for a rich answer you should construct a multiple regression model and comment on model coefficients, model fit and the statistical significance of results.

Format and submission

You are to work individually to complete these tasks. Your answers must be written in a minimum of font size 11 and your submission document must be no longer than 2 sides of A4 (including everything). The usual departmental penalties will be applied to overlength work. Don't forget to provide informative titles to your figures and also size them suitably.

The submission deadline is noon on Monday 13th December 2021. Submit the worksheet through Turnitin on the module Moodle page. Please make sure to enter your student exam candidate ID at the top of the page and also use it as your submission title on Moodle.

This worksheet counts for 10% of your GEOG0013 grade and will be marked using the same grading matrix used for the other worksheet tasks. Remember that 60% of your mark comes from the Field Notebook. Please see the Moodle guidance page (in the Week 9 section of the GEOG0013 page) for details of what to include in the notebook from the Data Analysis project.

Hints and tips

- Don't try and do everything! Part of the challenge is *selecting* appropriate techniques from the large number of methods introduced in the tutorials. This is not a tick-box exercise and there is no single 'right way' to tackle the tasks.
- It is helpful to think about *why* you expect variations in your x variables to be associated with variations in y. Formulating hypotheses about what you expect to find *before* running the analysis is a good move.
- Take care with writing style and the presentation of results. Try to write as concisely and precisely as possible.